

Segmentasi Perilaku Konsumsi Digital dan Dampaknya Terhadap Kesehatan Mental Menggunakan Algoritma K-Means

Dinda Apriliani Putri^{1*}, Albertus Radya Winny Putra², Devani Ula Ainurofiqi³,

Muhammad Afrizal Deviano Saputra⁴

¹Teknik Informatika

Universitas Duta Bangsa Surakarta

²Teknik Informatika

Universitas Duta Bangsa Surakarta

³Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}230103097@mhs.udb.ac.id

²230103086@mhs.udb.ac.id

³230103096@mhs.udb.ac.id

⁴Teknik Informatika

Universitas Duta Bangsa Surakarta

⁴230103108@mhs.udb.ac.id

Abstrak— Peningkatan pesat dalam penggunaan perangkat digital telah berdampak signifikan terhadap kesehatan mental masyarakat. Penelitian ini bertujuan untuk mengelompokkan perilaku konsumsi digital serta menganalisis dampaknya terhadap kesehatan mental menggunakan algoritma clustering K-Means. Kumpulan data yang digunakan berasal dari 2.000 responden dan mencakup 19 variabel numerik yang meliputi pola penggunaan perangkat digital, aktivitas digital, kondisi tidur, aspek psikologis, serta indikator kesehatan mental. Seluruh data dinormalisasi menggunakan metode Min-Max Scaling sebelum proses clustering dilakukan. Jumlah cluster optimal ditentukan menggunakan Elbow Method, yang menghasilkan nilai K=3. Proses clustering mengategorikan responden ke dalam tiga kelompok: Pengguna Sehat (Healthy User) sebanyak 31,85%, Pengguna Moderat (Moderate User) sebanyak 37,60%, dan Pengguna Berisiko Tinggi (High-Risk User) sebanyak 30,55%. Kualitas clustering dievaluasi menggunakan Silhouette Score dengan nilai sebesar 0,0627, yang tergolong lemah namun wajar untuk data perilaku manusia yang bersifat kontinum. Sistem ini diimplementasikan sebagai aplikasi web menggunakan PHP dan MySQL, serta menyediakan fitur untuk operasi CRUD, normalisasi data, clustering, dan evaluasi hasil.

Kata kunci— K-Means Clustering, Kesehatan Mental, Konsumsi Digital, Normalisasi Min-Max, Elbow Method, Silhouette Score

Abstract— The rapid increase in digital device usage has significantly impacted public mental health. This study aims to categorize digital consumption behaviors and analyze their impact on mental health using the K-Means clustering algorithm. The dataset comprises data from 2,000 respondents and includes 19 numerical variables covering digital device usage patterns, digital activities, sleep conditions, psychological aspects, and mental health indicators. All data were normalized using the Min-Max Scaling method prior to clustering. The optimal number of clusters was determined using the Elbow Method, resulting in K=3. The clustering process categorized respondents into three groups: Healthy Users (31.85%), Moderate Users (37.60%), and High-Risk Users (30.55%). Clustering quality was evaluated using the Silhouette Score, yielding a value of 0.0627; while this is considered low, it is reasonable given the nature of human behavioral data, which exists on a continuum. The system was implemented as a web application using PHP and MySQL, featuring capabilities for CRUD operations, data normalization, clustering, and result evaluation.

Keywords— K-Means Clustering, Mental Health, Digital Consumption, Min-Max Normalization, Elbow Method, Silhouette Score

I. PENDAHULUAN

Perkembangan teknologi digital yang pesat telah mengubah cara manusia berinteraksi, bekerja, dan menghabiskan waktu luang. Penggunaan perangkat digital seperti smartphone, laptop, tablet, dan televisi kini menjadi bagian tak terpisahkan dari kehidupan

sehari-hari. Indonesia memiliki lebih dari 167 juta pengguna media sosial, menjadikannya salah satu negara dengan populasi pengguna media sosial terbesar di dunia [1].

Namun di balik kemudahan yang ditawarkan, konsumsi digital yang berlebihan telah dikaitkan dengan berbagai masalah kesehatan mental,

seperti peningkatan tingkat kecemasan, depresi, gangguan tidur, penurunan mood, dan peningkatan stres. Penggunaan perangkat digital yang berlebihan telah dikaitkan dengan peningkatan masalah psikososial pada penggunanya [2]. Semakin tinggi intensitas penggunaan media sosial, semakin buruk kondisi kesehatan mental emosional yang dialami [3]. Tinjauan literatur terbaru turut menegaskan bahwa mahasiswa di era digital rentan mengalami stres, kecemasan, dan depresi akibat kecanduan media sosial, kelelahan digital, serta berkurangnya interaksi sosial nyata [4].

Untuk memahami pola hubungan tersebut secara sistematis, dibutuhkan metode analisis data yang mampu mengidentifikasi kelompok-kelompok responden berdasarkan kemiripan pola perilaku digitalnya tanpa memerlukan label kelas yang sudah diketahui sebelumnya. Salah satu pendekatan yang umum digunakan dalam data mining adalah algoritma K-Means Clustering, yaitu metode pengelompokan data yang bekerja dengan meminimalkan jarak antara data dengan centroid clusternya secara iteratif [5].

Kesehatan mental merupakan kondisi kesejahteraan psikologis yang sangat dipengaruhi oleh pola konsumsi teknologi digital sehari-hari [3]. Penggunaan media sosial yang berlebihan terbukti berkaitan erat dengan peningkatan risiko gangguan psikososial pada remaja, seperti kecemasan, depresi, dan penurunan kualitas tidur [1]. Untuk mengidentifikasi kelompok responden berdasarkan kemiripan pola perilaku tersebut, penelitian ini menggunakan algoritma K-Means Clustering, yaitu metode unsupervised learning yang bekerja dengan meminimalkan jarak Euclidean antara data dengan centroid clusternya secara iteratif [5]. Algoritma ini tetap menjadi metode clustering yang paling banyak digunakan karena kesederhanaan implementasi serta efisiensinya pada dataset berukuran besar [12]. Sebelum proses clustering, seluruh fitur dinormalisasi menggunakan Min-Max Scaling agar memiliki bobot yang setara dalam perhitungan jarak [6]. Jumlah cluster optimal (K) ditentukan menggunakan Elbow Method, yaitu teknik yang menghitung nilai Within-Cluster Sum of Squares (WCSS) pada setiap nilai K [7].

Kualitas hasil clustering selanjutnya dievaluasi menggunakan Silhouette Score, dengan rentang nilai -1 hingga +1, di mana nilai mendekati +1 menunjukkan cluster yang terpisah dengan baik [8]. Kombinasi Elbow Method dan Silhouette Score merupakan pendekatan evaluasi yang saling melengkapi dan banyak digunakan dalam berbagai studi clustering di Indonesia [9].

Berbagai penelitian terdahulu pada topik ini umumnya bersifat deskriptif atau korelasional dan berhenti pada identifikasi hubungan antara konsumsi digital dengan gejala kesehatan mental, tanpa mengelompokkan responden ke dalam segmen risiko yang terukur [1], [3], [10]. Penelitian yang telah menerapkan K-Means pada konteks serupa juga umumnya menggunakan dataset berskala kecil dan terbatas pada satu populasi tertentu, serta belum disertai evaluasi kualitas cluster secara kuantitatif maupun implementasi sistem yang dapat digunakan secara berkelanjutan [11]. Kesenjangan inilah yang coba dijawab oleh penelitian ini, yaitu dengan mengombinasikan segmentasi berbasis K-Means Clustering pada dataset berskala lebih besar (2.000 responden), evaluasi kualitas cluster menggunakan kombinasi Elbow Method dan Silhouette Score, serta implementasi hasil segmentasi ke dalam sistem berbasis web yang dapat digunakan untuk analisis lanjutan.

Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah: (1) bagaimana mengelompokkan responden berdasarkan pola perilaku konsumsi digital menggunakan algoritma K-Means Clustering; (2) bagaimana karakteristik masing-masing kelompok yang terbentuk dalam merepresentasikan dampaknya terhadap kesehatan mental; dan (3) bagaimana kualitas hasil pengelompokan tersebut apabila dievaluasi menggunakan Silhouette Score.

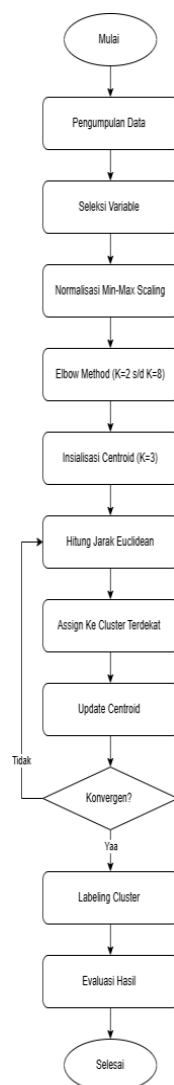
Penelitian ini bertujuan untuk mengelompokkan perilaku konsumsi digital responden menggunakan algoritma K-Means Clustering, menganalisis karakteristik tiap kelompok terhadap kesehatan mental, serta mengevaluasi kualitas hasil clustering menggunakan Silhouette Score. Hasil penelitian ini diimplementasikan dalam bentuk sistem

berbasis web menggunakan PHP dan MySQL, mencakup fitur manajemen data (CRUD), normalisasi Min-Max, proses clustering, Elbow Method untuk penentuan K optimal, visualisasi hasil, serta evaluasi kualitas clustering [8].

II. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Alur penelitian terdiri dari beberapa tahapan yang dilakukan secara berurutan, sebagaimana ditunjukkan pada flowchart berikut. Penjelasan tiap kotak pada flowchart adalah sebagai berikut



Gambar 6. Flowchart Alur Penelitian

B. Metode yang Digunakan

1) Normalisasi Min-Max

Normalisasi Min-Max adalah teknik preprocessing data yang mengubah nilai setiap fitur ke rentang [0, 1] menggunakan rumus:

$$X' = \frac{(X - X_{min})}{(X_{max} - X_{min})}$$

Normalisasi data merupakan langkah krusial dalam proses clustering untuk memastikan setiap fitur memiliki bobot yang setara dalam perhitungan jarak, terutama ketika fitur-fitur dalam dataset memiliki satuan dan skala yang berbeda-beda [6].

2) K-Means Clustering

K-Means adalah algoritma clustering yang termasuk dalam kategori unsupervised learning, bekerja dengan membagi dataset menjadi K kelompok berdasarkan kedekatan jarak antar data menggunakan jarak Euclidean:

$$d(x, c) = \sqrt{\sum (x_j - c_j)^2}$$

K-Means tetap menjadi algoritma clustering yang dominan di era big data karena kemampuannya beradaptasi dengan berbagai jenis dataset, termasuk data perilaku manusia yang berdimensi tinggi [5].

3) Elbow Method

Elbow Method digunakan untuk menentukan jumlah cluster optimal (K) dengan menghitung nilai WCSS (Within-Cluster Sum of Squares) untuk setiap nilai K [8]. Titik siku pada grafik WCSS terhadap K ditetapkan sebagai nilai K optimal karena penambahan cluster setelahnya tidak memberikan penurunan WCSS yang signifikan [7].

4) Silhouette Score

Silhouette Score digunakan untuk mengevaluasi kualitas clustering, dengan nilai berkisar antara -1 hingga +1; nilai mendekati +1 menunjukkan cluster yang terpisah dengan baik, sedangkan nilai negatif mengindikasikan data mungkin masuk ke cluster yang salah [12].

C. Arsitektur Sistem

Sistem direncanakan untuk diimplementasikan sebagai aplikasi web dengan komponen sebagai berikut:

- 1) Backend: PHP 8.x — logika K-Means, normalisasi, Elbow Method, dan evaluasi ditulis murni tanpa library eksternal.
- 2) Database: MySQL — menyimpan dataset, hasil clustering, dan log iterasi.
- 3) Frontend: HTML, CSS kustom, Bootstrap Icons, dan Chart.js untuk visualisasi grafik.
- 4) Server: Apache.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan dan Seleksi Dataset

Dataset yang digunakan diperoleh dari platform Kaggle dengan judul "*Impact of Screen Time on Mental Health*" yang dipublikasikan oleh Khushi Kyad, dapat diakses melalui tautan <https://www.kaggle.com/datasets/khushikyad001/impact-of-screen-time-on-mental-health>.

Dataset ini dipilih karena mencakup variabel yang komprehensif terkait pola penggunaan perangkat digital dan indikator kesehatan mental dalam satu kesatuan data, sehingga sesuai dengan kebutuhan penelitian ini tanpa memerlukan penggabungan (*merging*) dari beberapa sumber data yang berbeda.

Dataset yang digunakan adalah dataset Digital Diet & Mental Health yang terdiri dari 2.000 data responden dengan 25 atribut. Dari 25 atribut tersebut, dipilih 19 variabel numerik sebagai fitur clustering, sebagaimana ditunjukkan pada Tabel 1. Atribut kategorikal (*gender*, *location_type*, *uses_wellness_apps*, *eats_healthy*, *age*) dan identitas (*user_id*) tidak diikutsertakan dalam proses clustering karena tidak dapat dihitung jaraknya secara langsung dalam algoritma K-Means.

Tabel 1. Dataset

Variabel	Keterangan
<i>daily_screen_time_hours</i>	Total screen time per hari (jam)
<i>phone_usage_hours</i>	Durasi pemakaian HP per hari (jam)
<i>laptop_usage_hours</i>	Durasi pemakaian laptop per hari (jam)
<i>tablet_usage_hours</i>	Durasi pemakaian tablet per hari (jam)

Variabel	Keterangan
<i>tv_usage_hours</i>	Durasi pemakaian TV per hari (jam)
<i>social_media_hours</i>	Durasi penggunaan media sosial (jam)
<i>work_related_hours</i>	Durasi kerja/belajar digital (jam)
<i>entertainment_hours</i>	Durasi hiburan digital (jam)
<i>gaming_hours</i>	Durasi gaming per hari (jam)
<i>sleep_duration_hours</i>	Durasi tidur per hari (jam)
<i>sleep_quality</i>	Kualitas tidur (skala 1–10)
<i>mood_rating</i>	Rating mood harian (skala 1–10)
<i>stress_level</i>	Tingkat stres (skala 1–10)
<i>physical_activity_hours_per_week</i>	Aktivitas fisik per minggu (jam)
<i>mental_health_score</i>	Skor kesehatan mental (0–100)
<i>caffeine_intake_mg_per_day</i>	Konsumsi kafein per hari (mg)
<i>weekly_anxiety_score</i>	Skor kecemasan mingguan
<i>weekly_depression_score</i>	Skor depresi mingguan
<i>mindfulness_minutes_per_day</i>	Durasi mindfulness per hari (menit)

Dari 25 atribut yang tersedia, sebanyak 6 atribut dikeluarkan dari proses clustering karena algoritma K-Means berbasis jarak Euclidean hanya dapat memproses data numerik secara langsung tanpa transformasi tambahan seperti *one-hot encoding*. Sisa 19 atribut numerik dipertahankan seluruhnya tanpa reduksi maupun agregasi lebih lanjut. Keputusan ini didasarkan pada hasil pemeriksaan konsistensi data, di mana korelasi antara jumlah empat atribut durasi penggunaan perangkat (*phone_usage_hours*, *laptop_usage_hours*, *tablet_usage_hours*, *tv_usage_hours*) dengan atribut *daily_screen_time_hours* hanya sebesar 0,013, menunjukkan tidak adanya hubungan linear yang berarti antara keduanya. Selisih antara kedua nilai tersebut bervariasi dari -7,3 hingga +11,8 jam dengan rata-rata selisih 1,5 jam, dan sebanyak 34,2% dari 2.000 responden memiliki selisih lebih dari 3 jam. Temuan ini mengindikasikan bahwa *daily_screen_time_hours* kemungkinan dicatat secara independen (misalnya melalui estimasi subjektif responden) dan tidak merepresentasikan penjumlahan langsung durasi penggunaan per

perangkat, sehingga terdapat kemungkinan tumpang tindih waktu penggunaan antar perangkat (*concurrent usage*) atau perbedaan metode pencatatan pada tahap pengumpulan data. Oleh karena itu, agregasi atau rata-rata antar atribut durasi berisiko menghasilkan variabel turunan yang bias secara data. Penelitian ini juga tidak melakukan seleksi fitur lanjutan melalui analisis korelasi menyeluruh, karena seluruh 19 variabel secara konseptual merepresentasikan tiga dimensi utama yang ingin diukur; pola penggunaan perangkat digital, kondisi tidur, serta indikator psikologis/kesehatan mental sehingga eliminasi fitur dikhawatirkan menghilangkan informasi domain yang relevan. Analisis korelasi menyeluruh antar seluruh variabel disarankan sebagai langkah tambahan pada penelitian selanjutnya untuk mengevaluasi potensi multikolinearitas antar fitur.

B. Hasil Normalisasi Min-Max

Tahap preprocessing meliputi seleksi 19 variabel numerik yang relevan dan normalisasi Min-Max Scaling untuk menyamakan skala seluruh fitur ke rentang $[0, 1]$. Normalisasi Min-Max dipilih karena menghasilkan nilai dalam rentang yang terbatas dan tidak menghasilkan nilai negatif seperti pada metode Z-Score, sehingga lebih intuitif untuk data perilaku manusia [6]. Sebagai contoh, perbedaan skala signifikan antara *caffeine_intake* (0–400 mg) dan *sleep_quality* (1–10) berhasil disetarakan setelah proses normalisasi, sehingga tidak ada fitur yang mendominasi perhitungan jarak Euclidean pada tahap clustering berikutnya.

C. Hasil Penentuan Jumlah Cluster (Elbow Method)

Elbow Method dijalankan untuk $K=2$ hingga $K=8$. Nilai WCSS dihitung pada setiap iterasi K , dengan hasil ditampilkan pada Tabel 2.

Tabel 2. Penentuan Jumlah Cluster

Nilai K	WCSS	Penurunan WCSS	Keterangan
K=2	2052.09	—	—
K=3	1953.30	4.81%	Titik Elbow / Optimal
K=4	1883.63	3.57%	Mulai melambat
K=5	1829.35	2.88%	Melambat
K=6	1785.33	2.41%	Melambat
K=7	1745.65	2.22%	Melambat
K=8	1712.31	1.91%	Melambat

Berdasarkan Tabel 2, penurunan WCSS terbesar terjadi dari $K=2$ ke $K=3$ sebesar 4,81%. Setelah $K=3$, penurunan mulai melambat secara signifikan, sehingga $K=3$ ditetapkan sebagai jumlah cluster optimal. Hasil ini selaras dengan kategorisasi kondisi kesehatan mental secara domain yang umumnya dibagi dalam tiga tingkatan: baik, sedang, dan buruk.



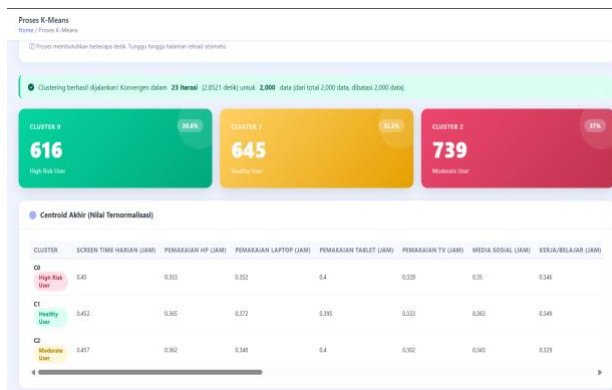
Gambar 7. Kurva Elbow Method

D. Hasil Proses K-Means Clustering

Proses K-Means dijalankan dengan parameter $K=3$, maksimal iterasi 100, dan metode inialisasi centroid Spread. Label cluster ditentukan secara otomatis berdasarkan rata-rata nilai *mental_health_score* setiap cluster: cluster dengan skor tertinggi diberi label Healthy User, skor sedang diberi label Moderate User, dan skor terendah diberi label High-Risk User. Pendekatan pelabelan otomatis berbasis domain ini sejalan dengan prinsip interpretabilitas hasil clustering[13].

Tabel 3. Distribusi Hasil Clustering

Cluster	Label	Jumlah Data	Persentase
0	High Risk User	616	30.80%
1	Healthy User	645	32.25%
2	Moderate User	739	36.95%
Total	—	2.000	100%



Gambar 8. Hasil Distribusi Clustering

Tabel 4. Profil Rata-rata Fitur per Cluster

Fitur	Healthy User	Moderate User	High Risk User
Screen Time (jam/hari)	6,01	6,08	5,98
Media Sosial (jam/hari)	2,10	1,99	2,03
Durasi Tidur (jam)	6,55	6,58	6,46
Kualitas Tidur (1–10)	5,54	5,86	5,25
Tingkat Stres (1–10)	7,24	2,50	7,41
Skor Kecemasan	9,85	9,43	10,47
Skor Depresi	11,25	9,58	9,36
Mental Health Score	52,08	49,43	47,37
Mindfulness (menit/hari)	10,26	11,17	10,77

Berdasarkan Tabel 4, tidak seluruh fitur menunjukkan pola perbedaan yang konsisten dan monoton antar ketiga cluster. Fitur *screen time*, durasi tidur, kualitas tidur, dan *mental_health_score* menunjukkan pola yang sejalan dengan urutan risiko (Healthy > Moderate > High-Risk atau sebaliknya), sedangkan fitur *stress_level* dan *weekly_depression_score* justru menunjukkan pola yang kurang konsisten — kelompok Healthy User memiliki rata-rata stres (7,24) yang mendekati kelompok High-Risk User (7,41), dan rata-rata skor depresi pada Healthy User (11,25) justru lebih tinggi dibandingkan Moderate User (9,58) maupun High-Risk User (9,36). Temuan ini mengindikasikan bahwa pelabelan cluster yang didasarkan pada satu variabel acuan (*mental_health_score*) tidak serta-merta mencerminkan keseragaman pada seluruh fitur psikologis lainnya, dan memperkuat interpretasi bahwa batas antar cluster pada data

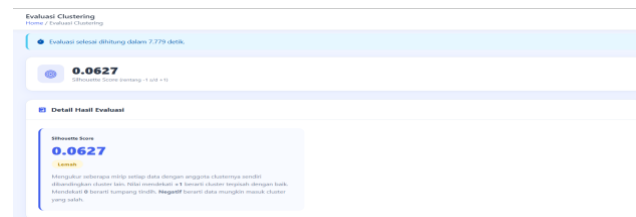
ini bersifat lunak (*soft boundary*), sebagaimana tercermin pula dari nilai Silhouette Score yang rendah (0,0627).

E. Hasil Evaluasi Kualitas Clustering

Evaluasi kualitas clustering dilakukan menggunakan Silhouette Score, dengan hasil ditampilkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Kualitas Clustering

Metrik	Nilai	Kategori	Keterangan
Silhouette Score	0.0627	Lemah	Wajar untuk data perilaku

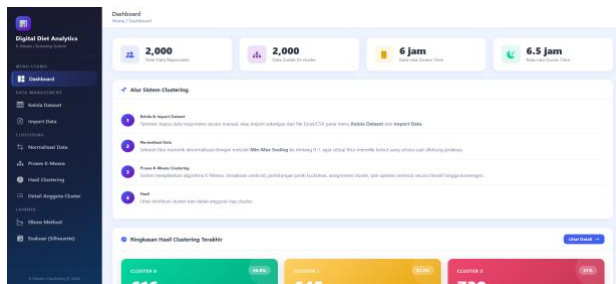


Gambar 9. Hasil Evaluasi Kualitas Clustering

Nilai Silhouette Score sebesar 0,0627 tergolong lemah, namun hal ini adalah kondisi yang wajar untuk dataset perilaku manusia. Data perilaku manusia bersifat kontinum dan tidak memiliki batas cluster yang tegas, sehingga nilai Silhouette Score rendah pada data jenis ini dapat diterima selama hasil clustering tetap bermakna secara domain[13]. Data yang berkaitan dengan kondisi psikologis manusia cenderung memiliki tumpang tindih antar kelas, yang tercermin dari nilai metrik evaluasi yang lebih rendah dibandingkan data teknis. Validitas hasil dalam penelitian ini lebih ditekankan pada interpretabilitas dan kesesuaian dengan konteks domain [14].

F. Hasil Implementasi Sistem

Sistem diimplementasikan dalam bentuk aplikasi web yang dapat diakses melalui browser. Tampilan sistem mencakup halaman dashboard, kelola dataset (CRUD), normalisasi data, Elbow Method, proses K-Means, hasil clustering, detail anggota cluster, dan evaluasi.



Gambar 10. Tampilan Dashboard Sistem

ID	Nama	Lokasi	Waktu	Status
1	Responden A	4.0 jam	10/10	OK
2	Responden B	4.0 jam	10/10	OK
3	Responden C	4.0 jam	10/10	OK
4	Responden D	4.0 jam	10/10	OK
5	Responden E	4.0 jam	10/10	OK
6	Responden F	4.0 jam	10/10	OK
7	Responden G	4.0 jam	10/10	OK
8	Responden H	4.0 jam	10/10	OK
9	Responden I	4.0 jam	10/10	OK

Gambar 11. Tampilan Kelola Dataset

Nama	Lokasi	Waktu	Status
Responden A	4.0 jam	10/10	OK
Responden B	4.0 jam	10/10	OK
Responden C	4.0 jam	10/10	OK
Responden D	4.0 jam	10/10	OK
Responden E	4.0 jam	10/10	OK
Responden F	4.0 jam	10/10	OK
Responden G	4.0 jam	10/10	OK
Responden H	4.0 jam	10/10	OK
Responden I	4.0 jam	10/10	OK

Gambar 12. Tampilan Normalisasi Data



Gambar 13. Tampilan Halaman Proses K-Means



Gambar 14. Tampilan Halaman Hasil Clustering

Secara keseluruhan, gambar-gambar yang ditampilkan pada bagian Hasil dan Pembahasan ini saling melengkapi untuk menggambarkan alur kerja sistem secara utuh: Gambar 2 memperlihatkan kurva Elbow Method yang menjadi dasar penentuan jumlah cluster optimal, Gambar 3 menampilkan distribusi hasil clustering ke dalam tiga kelompok responden,

Gambar 4 menunjukkan hasil evaluasi kualitas clustering melalui Silhouette Score, sedangkan Gambar 5 hingga Gambar 9 memperlihatkan antarmuka sistem berbasis web yang merepresentasikan tahapan penting dalam alur kerja sistem: halaman Dashboard (Gambar 5) berfungsi menyajikan ringkasan statistik dan visualisasi awal data responden; halaman Kelola Dataset (Gambar 6) menyediakan fungsi CRUD (create, read, update, delete) yang memungkinkan pengelolaan data responden secara mandiri; halaman Normalisasi Data (Gambar 7) menampilkan hasil transformasi Min-Max Scaling sebagai dasar penyetaraan skala antar variabel sebelum clustering dijalankan; halaman Proses K-Means (Gambar 8) memperlihatkan mekanisme iteratif perhitungan jarak dan pembaruan centroid hingga konvergen; serta halaman Hasil Clustering (Gambar 9) menyajikan keluaran akhir berupa label cluster untuk setiap responden. Tampilan lain yang secara fungsi maupun visual serupa, seperti formulir impor data, detail anggota cluster, serta halaman Elbow Method dan evaluasi yang telah direpresentasikan secara analitis pada Gambar 2 dan Gambar 4, tidak ditampilkan kembali agar pembahasan dapat difokuskan pada analisis hasil penelitian.

Jika ditinjau lebih lanjut, perbedaan karakteristik antar cluster pada Tabel 4 menunjukkan bahwa variabel screen time, durasi dan kualitas tidur, serta tingkat stres merupakan faktor yang paling konsisten membedakan ketiga kelompok, sedangkan skor mindfulness cenderung berbanding terbalik dengan tingkat risiko. Pola ini mengindikasikan adanya keterkaitan antara intensitas penggunaan perangkat digital dengan gangguan pola tidur dan peningkatan stres, yang pada akhirnya berkontribusi terhadap penurunan skor kesehatan mental, sejalan dengan temuan mengenai dampak penggunaan media digital terhadap gangguan psikososial [1] dan tinjauan literatur kesehatan mental mahasiswa di era digital [10].

Nilai Silhouette Score yang tergolong lemah (0,0627) perlu dimaknai sebagai konsekuensi dari sifat data perilaku manusia yang cenderung kontinum [13], bukan sebagai indikasi kegagalan

model. Nilai ini juga menyiratkan bahwa batas antar cluster, khususnya antara kelompok Moderate User dan kedua kelompok ekstrem, tidak sepenuhnya tegas, sehingga hasil pengelompokan lebih tepat dimanfaatkan sebagai pendekatan skrining awal dibandingkan sebagai diagnosis klinis yang bersifat pasti.

Dari sisi kontribusi praktis, sistem yang dibangun tidak hanya berfungsi sebagai alat visualisasi, tetapi juga berpotensi diterapkan oleh institusi pendidikan atau layanan konseling untuk melakukan skrining awal terhadap individu yang berpotensi masuk kategori High-Risk User, sejalan dengan pemanfaatan K-Means untuk segmentasi kesehatan mental mahasiswa berdasarkan faktor stres dan akademik [11] serta pendekatan berbasis EDA dan CRISP-DM untuk prediksi risiko media sosial terhadap kesehatan mental [4]. Dibandingkan dengan penelitian tersebut yang umumnya berhenti pada tahap analisis statistik, kontribusi utama penelitian ini terletak pada integrasi seluruh tahapan clustering, mulai dari normalisasi data hingga evaluasi, ke dalam satu sistem berbasis web yang dapat dioperasikan ulang oleh pengguna non-teknis. Meskipun demikian, sistem ini masih memiliki keterbatasan karena proses klusterisasi dijalankan secara statis dan belum mendukung pembaruan model secara otomatis ketika data responden baru ditambahkan, sehingga hal ini dapat menjadi arah pengembangan pada penelitian selanjutnya.

IV. KESIMPULAN

Penelitian ini berhasil membangun sistem clustering berbasis web menggunakan PHP dan MySQL yang mengintegrasikan seluruh tahapan analisis, mulai dari normalisasi Min-Max, penentuan jumlah cluster optimal melalui Elbow Method ($K=3$), proses K-Means, hingga evaluasi menggunakan Silhouette Score. Sistem ini berhasil mengelompokkan 2.000 responden ke dalam tiga cluster dengan karakteristik yang berbeda secara bermakna, yaitu Healthy User (31,85%), Moderate User (37,60%), dan High-Risk User (30,55%), yang terutama dibedakan oleh screen time, kualitas tidur, dan tingkat stres.

Kontribusi penelitian ini bagi bidang data mining terletak pada penerapan pendekatan

clustering yang tidak berhenti pada tahap analisis statistik, melainkan diintegrasikan ke dalam sebuah sistem yang dapat dioperasikan ulang secara mandiri, sehingga proses segmentasi dapat direplikasi pada dataset lain dengan struktur variabel serupa. Bagi bidang kesehatan mental, penelitian ini menawarkan pendekatan kuantitatif untuk mengidentifikasi kelompok berisiko berdasarkan pola konsumsi digital, yang dapat menjadi dasar bagi institusi pendidikan maupun layanan konseling dalam merancang intervensi yang lebih tepat sasaran, khususnya bagi kelompok High-Risk User.

Penelitian ini memiliki sejumlah keterbatasan. Pertama, analisis hanya menggunakan satu dataset dengan 2.000 responden yang belum tentu merepresentasikan populasi secara umum, sehingga generalisasi hasil perlu dilakukan dengan hati-hati. Kedua, nilai Silhouette Score yang diperoleh (0,0627) tergolong rendah dan mengindikasikan batas antar cluster belum sepenuhnya tegas, meskipun kondisi ini wajar terjadi pada data perilaku manusia yang bersifat kontinum. Ketiga, sistem yang dibangun masih bersifat statis dan belum mendukung pembaruan model secara otomatis ketika terdapat penambahan data responden baru.

Untuk penelitian selanjutnya, disarankan untuk menggunakan dataset dari sumber atau populasi yang lebih beragam guna menguji generalisasi hasil, mencoba metode clustering lain seperti DBSCAN atau Fuzzy C-Means yang lebih toleran terhadap data dengan batas cluster tidak tegas [9], [15], serta melakukan seleksi fitur untuk mengidentifikasi variabel yang paling berpengaruh terhadap pengelompokan perilaku digital dan kesehatan mental.

REFERENSI

- [1] I. A. N. Gunawan, S., and I. Shalahuddin, "Dampak Penggunaan Media Sosial Terhadap Gangguan Psikososial Pada Remaja: A Narrative Review," *Jurnal Kesehatan*, vol. 15, no. 1, pp. 78–92, Jun. 2022, doi: 10.23917/jk.v15i1.17426.
- [2] M. Dorrestein, S. B. Nutley, and L. B. Thorell, "Screen Time, Addictive Use of Social Media, Motives for Social Media Use and Social Media Content: Interrelations and Associations with Psychosocial Problems," *Int. J. Ment. Health Addict.*, Jun. 2025, doi: 10.1007/s11469-025-01491-5.

- [3] J. Kesehatan Mercusuar, U. Suryani, Y. Okjevi, F. Kesehatan dan Sains, and U. Mercubaktijaya, "HUBUNGAN INTENSITAS PENGGUNAAN MEDIA SOSIAL TERHADAP KESEHATAN MENTAL EMOSIONAL REMAJA", [Online]. Available: <http://jurnal.mercubaktijaya.ac.id/index.php/mercusuar>
- [4] F. Ayuning Tyas, A. Basir, A. Elistya Ardhin, and P. Korespondensi, "PREDIKSI RESIKO PENGGUNAAN MEDIA SOSIAL TERHADAP KESEHATAN MENTAL MENGGUNAKAN EXPLORATORY DATA ANALYSIS (EDA) DAN CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)," vol. 13, no. 2, pp. 487–498, 2026.
- [5] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (N. Y.)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/J.INS.2022.11.139.
- [6] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 330–338, Sep. 2022, doi: 10.47065/josyc.v3i4.2112.
- [7] N. A. Maori, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," *Jurnal SIMETRIS*, vol. 14, 2023.
- [8] M. Sholeh and K. Aeni, "PERBANDINGAN EVALUASI METODE DAVIES BOULDIN, ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN MENGGUNAKAN ALGORITMA K MEANS." [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Trave>
- [9] A. Saputra and R. Yusuf, "Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1346–1361, Jul. 2024, doi: 10.57152/malcom.v4i4.1516.
- [10] B. Jurnal, E. Karyati, E. Agustina Julisawati, and J. Selatan, "Kesehatan Mental Mahasiswa di Era Digital: Tinjauan Literatur Psikologi Klinis dan Sosial," 2025, doi: 10.63822/pmx2a928.
- [11] J. Parhusip, J. Feranita, M. Pradana Mentari, T. Apriliani, and A. Marcello Jessend, "Implementasi Algoritma K-Means Clustering Untuk Segmentasi Tingkat Kesehatan Mental Mahasiswa Berdasarkan Faktor Stress dan Akademik," 2026.
- [12] D. Chicco, A. Campagner, A. Spagnolo, D. Ciucci, and G. Jurman, "The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters," *PeerJ Comput. Sci.*, vol. 11, p. e3309, Nov. 2025, doi: 10.7717/peerj-cs.3309.
- [13] C. X. Gao *et al.*, "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Res.*, vol. 327, p. 115265, Sep. 2023, doi: 10.1016/J.PSYCHRES.2023.115265.
- [14] D. Rifaldi, Abdul Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, Apr. 2023, doi: 10.51454/decode.v3i2.131.
- [15] I. H. Zahro, U. A. Rosyidah, and L. Handayani, "Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Provinsi di Indonesia Berdasarkan Kualitas Perguruan Tinggi," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 1, pp. 80–86, Feb. 2024, doi: 10.37148/bios.v5i1.102.